

4. А.Н. Важдаев Технология создания информационных систем в среде 1С: Предприятие: учебное пособие / А.Н. Важдаев. – Юрга: Издательство Юргинского технологического института (филиал) Томского политехнического университета, 2007. – 132с.

## СОЗДАНИЕ ЕДИНОГО КОМПЛЕКСА ЭФФЕКТИВНЫХ ИНСТРУМЕНТОВ СЕТЕВОГО АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

*А.И. Труфанов<sup>1</sup>, А.Ф. Тухватуллина<sup>1</sup>, И.А. Лызин<sup>2</sup>, М.А. Тараник<sup>2</sup>*

<sup>1</sup> (Иркутск, Иркутский национальный исследовательский  
технический университет)

*troufan@gmail.com*

<sup>2</sup> (Томск, Томский политехнический университет)

*i-lyzin@mail.ru, taranik@tpu.ru*

## THE ESTABLISHMENT OF A UNIFIED SET OF EFFECTIVE TOOLS NETWORK TIME SERIES ANALYSIS

*A.I. Trufanov<sup>1</sup>, A.F. Tukhvatullina<sup>1</sup>, I.A. Lyzin<sup>2</sup>, M.A. Taranik<sup>2</sup>*

<sup>1</sup> (Irkutsk, Irkutsk national research technical university)

<sup>2</sup> (Tomsk, Tomsk Polytechnic University)

**Annotation.** This article presents methods and algorithms to create a unified set of effective tools network time-series analysis. The relevance of the work is underpinned to the fact that network analysis is a modern method of research Big Data in time series format and pictures. The results may be used to solve tasks in many areas of application.

**Key words:** BigData, bronchial asthma, algorithm, cluster analysis.

**Введение.** Общее накопленное человечеством к настоящему моменту количество информации отдельный индивид не в состоянии переработать, осознать, понять. Поэтому существует необходимость свести данные к более понятным и простым, определяя и измеряя их показатели – метрики. Подобные операции достаточно сложно провести при анализе данных, представленных в виде временных рядов. BigData – это обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия [1]. Одним из эффективных и удобных современных способов обработки BigData является их трансформация в сети.

Концепцию преобразования таких данных в формате временных рядов в комплексные сети впервые предложила группа ученых, в составе: Майкл Смол (Michael Small), Цзе Чжан (Jie Zhang) и Сяоке Сюй (Xiaoke Xu) в 2006г.. Авторы при преобразовании данных в сеть акцентировали внимание прежде всего на длине пути и кластеризации узлов внутри сети [2].

Этот подход к анализу больших данных начинает использоваться во многих сферах, в том числе таких, как: машиностроение, здравоохранение, экономика, банковское дело, геология и так далее.

Разработка высокопроизводительных алгоритмов имеет большое значение для исследований многих сфер жизни, так как они помогают существенно снизить затраты времени на обработку больших данных.

К данному моменту предлагаются такие алгоритмы конвертирования временных рядов в сети, как: алгоритм прямой видимости, алгоритм горизонтальной видимости, параметрический алгоритм и ряд других.

В настоящее время не существует заявленного единого набора удобных и практичных инструментов сетевого анализа временных рядов. То есть другими словами, разработчиками не предлагается комплекс приложений, на вход которого подаётся график временного ряда, а на выходе получается вектор сетевых метрик.

Цель работы заключается в создании единого комплекса эффективных инструментов сетевого анализа временных рядов. Для достижения цели предполагается решение следующего перечня задач:

1) Необходимо извлечь координаты точек из изображений графиков, подобрав и настроив подходящее приложение. Для этого из перечня программ по оцифровке графиков путём исследования будет выбрано в достаточной мере точное и удобное.

2) Используя современный язык программирования, требуется разработать инструмент трансформации данных в сеть, способный установить связи между узлами с помощью известных алгоритмов.

3) Визуализировать сеть и проанализировать её метрики, с помощью бесплатного, простого, доступного и производительного приложения.

4) Выбрать необходимое приложение кластерного анализа и произвести необходимые расчёты.

**Предметная область.** Предметом исследования являются графические данные исследования особенностей различных форм бронхиальной астмы. Бронхиальная астма (от греческого *asthma* – тяжелое дыхание, удушье) – это хроническое заболевание лёгких, поражающее людей всех возрастных групп. Оно может протекать в виде единичных, эпизодических приступов либо иметь тяжёлое течение с астматическим статусом и летальным исходом. По данным медицинской статистики, за последние годы заболеваемость бронхиальной астмой в большинстве стран значительно возросла. Увеличение распространенности заболевания среди лиц молодого возраста указывает на сохраняющуюся тенденцию роста частоты этого заболевания. Несмотря на научные достижения в области этиологии и наличие новых лекарственных средств, заболеваемость и смертность от бронхиальной астмы постоянно возрастают. Это характерно для большинства стран мира [3].

Сотрудниками Сибирского государственного медицинского университета и Томского политехнического университета на протяжении нескольких лет проводились исследования обнаружению достоверных различий между группами пациентов (по физиологическим и психологическим показателям) с заболеванием бронхиальной астмой, разделенных по степени влияния психосоциальных факторов на возникновение, развитие и течение болезни. Для выявления скрытых закономерностей в экспериментальных данных были использованы методы структурного анализа данных, в том числе, методы научной визуализации [4].

Экспериментальные данные представляли собой кривые дыхания у пациентов с различными типами бронхиальной астмы и у группы условно здоровых людей, зарегистрированных с помощью прибора «MONITOR». Данный прибор фиксировал значения кривой дыхания с частотой дискретизации 6 Гц. Показатели снимались как у пациентов с бронхиальной астмой, так и у здоровых людей на протяжении 3 часов в ночное время суток.

Кривые дыхания можно представить, как последовательность значений, взятых в дискретные моменты времени  $t_i$ . Промежутки времени между последовательными отсчётами (интервалы дискретизации)  $\Delta t_i = t_i - t_{i-1}$  постоянны и в нашем случае равны.

**Методы.** Первым этапом требуемого решения задачи является оцифровка графика.

Под оцифровкой документов понимают процесс перевода документов из традиционной (бумажной) формы в электронный (цифровой) вид с помощью специальных технических средств. Вместо термина «оцифровка» иногда употребляют схожее понятие – сканирование.

Зачастую исходные данные, по которым построен график, не публикуются, а представляются в виде графиков или кривых. Отсюда встает вопрос о получении координат, по которым построен график. Изобилие современных программных продуктов для оцифровки графиков зачастую не удовлетворяет требованиям пользователей. Например, сложный интерфейс, некорректность вычислительных данных, невозможность обработки изображений плохого качества, отсутствие инструментальных средств для более точных вычислений [5].

Основные требования к программам оцифровки графиков:

- 1) малая погрешность вычислений;
- 2) дружелюбный интерфейс;
- 3) возможность обрабатывать любые форматы изображений;
- 4) возможность обрабатывать изображения плохого качества.
- 5) наличие как автоматического, так и ручного режима работы.

После проведения серии экспериментов на различных отсканированных графиках, были получены следующие результаты: программа Engauge Digitizer 4.1. показала самые точные данные оцифровки не только в автоматическом, но и в ручном режиме, поэтому была включена в состав программных средств, используемых в работе.

Второй этап работы заключается в реализации основных алгоритмов конвертирования данных в сеть. Ввиду отсутствия в открытом доступе бесплатных и удобных программных средств, подходящих для выполнения данного этапа работы, было принято решение самостоятельно создать такой инструмент. Кроме того, необходим инструмент, позволяющий развивать методы, алгоритмы и схемы, что невозможно для имеющихся готовых, стандартных приложений. В качестве среды разработки использовались средства Microsoft Visual Studio, а язык программирования C#.

Сеть считается сформированной тогда, когда установлены связи между её элементами. На этом этапе работы необходимо построить связи между точками и получить матрицу смежности  $A$ , где значение  $a_{ij} = 1$  будет означать, что связь между элементами  $i$  и  $j$  есть, 0 – связь отсутствует.

Можно выделить два класса решаемых задач:

- 1) исследование эффективности и уязвимости, определяемых топологией систем;
- 2) сетевые приемы распознавания свойств систем и угроз для нее.

Первый класс задач предполагает: построение сетевой онтологии рассматриваемой системы, конструирование соответствующей реальной сети или синтетической сети, максимально приближенной к реальной, определение сетевых метрик, отвечающих за топологическую эффективность и стойкость, разработку возможных стратегий угроз, имитация атак на сетевую структуру и оценка сопутствующего ущерба, выяснение уязвимых мест сети, нахождение путей оптимизации системы. Решение задач данного класса позволяет приблизиться к построению изначально структурно-производительной и устойчивой системы.

Второй класс задач связан с анализом данных, представляемых в виде временных рядов или изображений. Безусловно, существуют и иные методы анализа, однако конвертирование данных в сеть отличается тем, что позволяет решать многие объемные задачи быстро и красиво.

Характерно, что анализ одномерных (или многомерных) временных рядов с помощью комплексных сетей недавно привлек внимание исследователей, работающих в широком диапазоне предметных областей [5]. При этом за последнее десятилетие был предложен ряд методологий для преобразования одномерных и многомерных временных рядов в комплексные сети.

Конвертирование временного ряда с построением связей между узлами будет осуществлено с помощью двух алгоритмов прямой (NVG) и горизонтальной (HVG) видимости.

Алгоритмы видимости – это семейство правил для отображения временных рядов реальных значений на сети. Во всех этих алгоритмах каждому значению временного ряда ставится в соответствие узел, но критерий соединения узлов связью в алгоритмах отличается.

Алгоритм горизонтальной видимости (HVG) схож с алгоритмом прямой видимости (NVG), однако для построения связей между точками строится прямая строго параллельная оси  $OX$ .

Немаловажным этапом при анализе будет визуализация данных так как это один из лучших способов донесения нужной информации, которая сразу же привлекает внимание к ключевым моментам контента. В общем случае, инструменты визуализации графов делятся на две категории – это программы, позволяющие строить графы вручную (специализирован-

ные редакторы) и программы, оснащенные инструментарием и интерфейсом для автоматизированной визуализации графа на основании введенных данных. Конечно же, существует и третий вариант – многие современные языки и среды программирования имеют в распоряжении библиотеки, пользуясь которыми, можно создать пользовательский интерфейс, отвечающий определенным требованиям. В работе было принято решение остановиться на уже готовом инструментарии, который на данный момент доступен специалистам по большим данным.

Среди приложений, осуществляющих построение графов в автоматическом режиме, Gephi является самой эффективной и удобной программой. В связи с этим в данной выпускной квалификационной работе для анализа сети мы будем использовать именно этот программный продукт.

И наконец завершающим этапом работы является выполнение кластерного анализа. Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя.

Спектр применений кластерного анализа очень широк: его используют в археологии, медицине, психологии, химии, биологии, государственном управлении, филологии, антропологии, маркетинге, социологии, геологии и других дисциплинах. Однако универсальность применения привела к появлению большого количества несовместимых терминов, методов и подходов, затрудняющих однозначное использование и непротиворечивую интерпретацию кластерного анализа.

Для данного этапа работы была выбрана программа Clusters 5.10 для MS-Windows. Так как именно это приложение в полной мере сочетает в себе простоту и функциональность необходимую для данного этапа.

**Заключение.** В ходе выполнения данной работы были решены поставленные задачи по созданию единого комплекса эффективных инструментов сетевого анализа временных рядов.

Этот набор инструментов включает в себя элементы:

- 1) оцифровки графических данных (адаптированное приложение Enguage);
- 2) трансформации временных рядов в цифровой формате в комплексную сеть (оригинальное приложение, реализующее NVG, HVG алгоритмы);
- 3) сетевого анализа и визуализации (адаптированное приложение Gephi);
- 4) кластерного анализа (адаптированное приложение Clusters 30).

Дальнейшая работа будет посвящена разработке собственного приложения для трансформации временных рядов и тестированию комплекса инструментов сетевого анализа на примере графических данных исследования особенностей различных форм бронхиальной астмы.

**Исследование выполнено при частичной финансовой поддержке РФФИ в рамках выполнения проекта 18-07-00543.**

## ЛИТЕРАТУРА

1. Большие данные (Big Data) – tadviser [Электронный ресурс]. – Режим доступа: [http://www.tadviser.ru/index.php/Статья:Большие\\_данные\\_\(Big\\_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)).
2. J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. Physical Review Letters, 96:238701, 2006.
3. Берестнева О.Г., Немеров Е.В., Степанов Д.Ю., Шевелев Г.Е., Осадчая И.А.// Исследование компьютерной визуализации в анализе кривых дыхания (на примере исследования особенностей различных форм бронхиальной астмы). – Институт кибернетики Нацио-

нального исследовательского Томского политехнического университета, 2Сибирский государственный медицинский университет.

4. Osadchaya I. A., Berestneva O. G., Volovodenko V. A., Marukhina O. V. Multidimensional Data Visualization Methods Based on Generalized Graphic Images // Communications in Computer and Information Science. — 2014. — Vol. 466. — p. 568-576.

5. Сравнительный анализ программ оцифровки графиков – Электронный научно-практический журнал «Современные научные исследования и инновации» [Электронный ресурс]. – Режим доступа: <http://web.snauka.ru/issues/2015/11/60095>.

## ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ АНАЛИЗА ДАННЫХ ЭЛЕКТРОНЕЙРОМИОГРАФИЧЕСКИХ ИССЛЕДОВАНИЙ

*В.А. Шамаков\*, О.Г. Берестнева, Е.И. Губин, С.А. Рудченко  
(г. Томск, Томский государственный университет)  
e-mail: sva1.0@mail.ru*

## INFORMATION TECHNOLOGY OF DATA ANALYSIS OF NEUROMYOGRAPHIC RESEARCHES

*V.A. Shamakov\*, O.G. Berestneva, E.I. Gubin, S.A. Rudchenko  
(Tomsk, Tomsk State University)  
e-mail: sva1.0@mail.ru*

**Abstract.** Currently, neurological diseases and, in particular, various kinds of polyneuropathy are one of the urgent problems of mankind. The paper presents the preliminary results on the automation of the process of formation of medical opinion on the basis of the analysis of electroneuromyographic studies. Also shows "raw" data and prepared conclusions and electroneuromyograms. The data was preprocessed using the Python programming language. All work is carried out on the data of real electroneuromyographic studies.

**Keywords:** polyneuropathy, data science, data mining, machine learning, IT, ENMG

В настоящее время число неврологических заболеваний растет, и эта проблема является весьма актуальной в нашем современном мире[1].

*Электронеиromиография (ЭНMG) – это комплексное электрофизиологическое исследование, необходимое для определения функционального состояния периферической нервной системы и мышц. Методика позволяет выявлять патологические изменения на самых ранних стадиях заболеваний. Электронеиromиография позволяет регистрировать и анализировать параметры вызванных потенциалов (ВП) мышцы и нерва (латентный период, форма, амплитуда и длительность ВП); определять число функционирующих двигательных единиц (ДЕ), а также скорость проведения импульса (СПИ) по двигательным и чувствительным волокнам периферических нервов. Проводить подсчет мотосенсорного и краниокаудального коэффициентов, коэффициентов асимметрии и отклонения от нормы.*

Показания к ЭНMG следующие: туннельные синдромы, невриты, радикулопатия (при остеохондрозе и болях в спине), плексопатии, сахарный диабет (для выявления полинейропатии нижних и верхних конечностей), БАС, болезнь мотонейрона.

Медицинские центры с оборудованием для проведения ЭНMG находятся в большинстве крупных городов России и мира.

В Томске данными исследованиями занимается «Лаборатория нейрофизиологии и функциональной диагностики». Лаборатория проводит множество видов исследований, многие из которых входят в состав общего медицинского страхования. За последние 5 лет в клинике прошло обследование около 6000 человек. Наиболее часто встречающимся видом заболевания является полинейропатия.